

## **SyGMA: combining expert knowledge and empirical scoring in the prediction of metabolites**

L. Ridder

M. Wagener

*Molecular Design & Informatics Department, Organon, part of Schering-Plough, 5340 BH Oss,  
The Netherlands*

Predictions of potential metabolites, based on chemical structure, are becoming increasingly important in drug discovery to guide medicinal chemistry addressing metabolic issues and to support experimental metabolite screening and identification. We present a novel rule-based method, SyGMA (Systematic Generation of Metabolites), to predict potential metabolites of a given parent structure. A set of reaction rules covering a broad range of phase 1 and phase 2 metabolism has been derived from metabolic reactions reported in the Metabolite database to occur in man. An empirical probability score is assigned to each rule representing the fraction of correctly predicted metabolites in the training database. This score is used to refine the rules and to rank predicted metabolites. To obtain a better overview of which metabolic reactions are reproduced / not reproduced by SyGMA, and to support ongoing efforts to extend the rules, a similarity analysis of the reactions present in the database was performed and mapped with the SyGMA results. The current rule set of SyGMA covers approximately 70% of metabolic reactions observed in man.

Evaluation of the rule based predictions demonstrated a significant enrichment of true metabolites in the top of the ranking list: while, in total, 68% of all observed metabolites in an independent test set were reproduced by SyGMA, a large part, 30% of the observed metabolites, were identified among the top 3 predictions. From a subset of cytochrome P450 specific metabolites, overall 84% were reproduced, with 66% in the top 3 predicted phase 1 metabolites.

Specific examples are given to demonstrate the usage of SyGMA in experimental metabolite identification as well as the application of SyGMA to suggest chemical modifications improving the metabolic stability of compounds.

Finally, a method will be presented to supplement each prediction with the most similar experimental examples in a metabolite database. This enables the user to quickly assess with more detail the value of individual SyGMA predictions.